Article:

Thornton, Michael Alan orcid.org/0000-0002-4470-809X (2019) Exact Discrete Representations of Linear Continuous Time Models with Mixed Frequency Data. Journal of Time Series Analysis. 5. pp. 951-967. ISSN 1467-9892

https://doi.org/10.1111/jtsa.12471

Exact Discrete Representations of Linear Continuous Time Models with Mixed Frequency Data

Michael A. Thornton
University of York
April 2019

Abstract

The time aggregation of vector linear processes containing (i) mixed stock-flow data and (ii) aggregated at mixed frequencies, is explored, focusing on a method to translate the parameters of the underlying continuous time model into those of an equivalent model of the observed data. Based on manipulations of a general state-space form, the results may be used to model multiple frequencies or aggregation schemes. Estimation of the continuous time parameters via the ARMA representation of the observable data vector is discussed and demonstrated in an application to model stock price and dividend data. Simulation evidence suggests that these estimators have superior properties to the traditional approach of concentrating the data to a single low frequency.

Keywords. Time aggregation; CARMA process; mixed frequency; state space; the exact discrete representation.

J.E.L. classification number. C32 **Acknowledgements**: I am grateful to the editor and two referees for careful reading and for other suggestions which improved the paper as well as to participants of the Peter C.B. Phillips honorary conference at the University of York, the Mixed Frequency Econometrics conference at the University of Essex, the conference on Computational and Financial Econometrics 2017 and at the Agence Français de Devéloppement, particularly Florent McIsaac, for helpful comments. All errors remain my own.

Address for Correspondence: Michael A. Thornton, Department of Economics and Related Studies, University of York, Heslington, York, YO10 5DD, England.
Tel: +44 1904 324566; fax: +44 1904 ; e-mail: michael.thornton@york.ac.uk.

1 Introduction

The fact that many important macroeconomic series are available only quarterly, while others may be available monthly and some financial series are available at a daily or higher frequency has led to a growing interest among multivariate time series analysts in techniques designed to model data available at different frequencies. The traditional response, to express the data at a common low frequency through aggregation or systematic sampling, is not only seen as a wasteful use of data but also bypasses questions of significant practical importance, such as how one should model the current value of a low frequency variable given more recent observations of high frequency covariates, see Foroni and Marcellino (2013) for an overview.

A variety of methods have been proposed to build models that utilise mixed frequency data efficiently, with an evident split between approaches that feature an underlying model at some fundamental frequency and those that do not. Prominent among the latter group is the Mixed Data Sampling (MIDAS) approach of Ghysels et al. (2004) and Ghysels (2016), in which the mass-parameterisation problem that results from regressing a low frequency variable on a large number of lags of a number of high frequency series is overcome by the use of constrained lag polynomials. Where that problem is less pronounced, the unrestricted version, or U-MIDAS, of Foroni et al. (2015) may be applied. The comparison of MIDAS with an alternative method is given in Schumacher (2016) in which bridge equations are used to interpolate a high frequency model of the low frequency variables.

More fundamental approaches, which attempt to estimate a set of deep parameters for a single frequency model using mixed frequency data, have their origins in the multivariate temporal aggregation studies of, among others: Lütkepohl (1987); Marcellino (1996) and Marcellino (1999). Techniques include the use of the Kalman filter, see Zadrozny (1988) and Seong et al. (2013) and an adaptation of the Yule Walker equations, see Chen and Zadrozny (1998) and Zadrozny (2016). This approach has the advantage of enabling the analyst to draw inference or to impose *a priori* restrictions on the parameters of the fundamental model. In most cases, the fundamental model is phrased in discrete time, although Zadrozny (1988) is a notable exception.

There is a long and venerable tradition of macroeconomic modelling in continuous time, much of it following the pioneering work of Rex Bergstrom, see Bergstrom (1990) and Bergstrom and Nowman (2007). This approach has been based on the construction and evaluation against data of the exact discrete representation: a mapping from the continuous time parameters to the first and second order properties of a corresponding discretely observed time series. In addition to being of interest in its own right, the exact discrete time model is computationally more efficient than the Kalman filter approach, once the set-up costs of deriving the discrete time model have been borne, Bergstrom (1985). That tradition has, however, tended to ignore the possibility of mixed frequency data. Despite the availability of, for example, monthly observations for some series, the model of Bergstrom and Nowman (2007) is quarterly. Recently Chambers (2016) has sought to address this issue, providing results for the first order continuous time model. This paper generalises those results to the second and higher order models often used in applied work.

We derive a mapping from the continuous time parameters to the first and second order properties of a corresponding mixed frequency time series, which we refer to as the exact discrete representation for mixed frequency data. We allow the time series to contain both systematically sampled *stock* variables, such as prices or interest rates, and time aggregated *flow* variables, such as GDP or profits. The vectors of high and of low frequency observations are allowed to contain data of both types, generalising the results in Chambers (2016) in which the high frequency variables were stocks and the low frequency flows. This is sufficiently general to cover a wide range of data sources. The case in which a (non-uniform) weighted average of data collected over different points of the observation cycle would require more expansive methods.

It is common to conceptualise a model of temporal aggregation or mixed frequency data in discrete time as a matter of missing data in a larger state space model. That approach is not open when the missing data forms a continuous record and so other methods have been deployed in, for example, Chambers and Thornton (2012), Thornton and Chambers (2017), based around: casting the model in state-space; performing the aggregation; before finally, solving out for the ARMA representation. This paper follows that approach. Not only can this be extended to mixed frequency data, it can be iterated to cover multiple frequencies. In contrast to a missing observations approach, the number of unobservable elements of the state vector will not increase as a result of the mixed frequency aggregation, but will remain determined by the autoregressive order of the continuous time system and the number of flow variables. This is likely to offer computational advantages when the high frequency variables are available with much higher frequency than the low.

Section 2 covers the journey from a continuous time ARMA model to an analogue describing a mixed frequency vector. Section 3 explains the estimation algorithm and section 4 includes a simulation study, suggesting that a mixed frequency technique out-performs traditional estimation methods based on the concentration of data to a single frequency. Section 5 contains an application of the methods to a well-known data set of (high frequency) stock price and (low frequency flow) dividend data, and section 6 concludes.

In such a treatment it can be difficult to serve those competing masters: generality and comprehensibility. While all notation is explained when introduced in the text, the following is a short guide to the conventions used in the paper. In general, $x(t)$ and x_t denote a sequence of vectors of variables of interest in continuous and discrete time respectively, ξ_t a sequence of state vectors, typically including x_t alongside unobservable elements. The scalar n denotes a dimension. The superscripts f refers to flow variables and s to stocks and for mixed frequencies, h denotes high and l denotes low, appearing as superscripts on state (sub-)vectors and subscripts when partitioning matrices. Superscripts in parentheses refer to the level of aggregation. The symbol $\tilde{.}$ is generally used when a system has been augmented to aggregate flows. The expression $\lfloor . \rfloor$ denotes the largest integer less than or equal to while $\lceil . \rceil$ denotes the smallest integer greater than or equal to.

2 Mixed frequency CARMA

2.1 Model

The continuous time ARMA (p, q) model for the $n \times 1$ vector $x(t)$ is given by

$$\begin{aligned} D^p x(t) &= a_0 + A_{p-1} D^{p-1} x(t) + \ldots + A_0 x(t) \\ &\quad + e(t) + V_1 De(t) + \ldots + V_q D^q e(t), \; t > 0, \end{aligned} \tag{1}$$

where D denotes the mean square differential operator [1]. The $n \times 1$ vector $e(t)$ is a continuous time white noise process with zero mean and second order properties

$$E\left[\int_{t_1}^{t_2} e(r)dr \int_{t_1}^{t_2} e(s)'ds\right] = \Sigma\left(t_2 - t_1\right),$$

$$E\left[\int_{t_1}^{t_2} e(r)dr \int_{t_1}^{t_2} e(\tau + s)'ds\right] = 0, \quad |\tau| > |t_2 - t_1|.$$

The condition that $p > q$ is important in ensuring that $x(t)$ has an integrable spectral density matrix and hence a finite variance. The parameters of interest are the $n \times 1$ vector, a_0, and $n \times n$ matrices $A_0, \ldots, A_{p-1}, V_1, \ldots, V_q$ and finally the symmetric matrix Σ.

For these parameters to be estimated without access to a continuous record we write (1) as a first order differential equation. In keeping with Zadrozny(1988), Chambers and Thornton (2012) and Thornton and Chambers (2017), define the $np \times 1$ state vector $y(t) = [x(t)', y_2(t)', \ldots, y_p(t)']'$. Then equation (1) may be expressed in state space form

$$Dy(t) = a + Ay(t) + Ve(t), \tag{2}$$

where

$$A = \begin{pmatrix} A_{p-1} & I & 0 & \ldots & 0 \\ A_{p-2} & 0 & I & \ldots & 0 \\ \vdots & & & & \vdots \\ A_1 & 0 & 0 & \ldots & I \\ A_0 & 0 & 0 & \ldots & 0 \end{pmatrix}, \quad V = \begin{pmatrix} V_{p-1} \\ V_{p-2} \\ \vdots \\ V_1 \\ I \end{pmatrix}, \quad a = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ a_0 \end{pmatrix},$$

with $V_j = 0$ for $j > q$. This is not the only possible representation of the system but the capacity to hold derivatives of $e(t)$ give it preference over rivals, see discussion in Thornton and Chambers (2017).

The task is to solve (1) in such a way that it produces a law of motion for an observable data vector. We imagine that $x(t)$ contains both *stock* variables, which are systematically sampled, and *flow* variables, which are time aggregates, and that variables of each type feature in the high and low frequency data vectors. Without loss of generality we first

[1] More precisely, if $x(t)$ is mean square differentiable then there exists a process $\xi(t)$ satisfying

$$\lim_{\delta \to 0} E\left\{\frac{x(t+\delta) - x(t)}{\delta} - \xi(t)\right\}^2 = 0,$$

which we denote $Dx(t) = \xi(t)$

partition the vector $x(t)$ as

$$x(t)=\begin{bmatrix} x^{s}(t) \\ \dots \\ x^{f}(t) \end{bmatrix}=\begin{bmatrix} x^{hs}(t) \\ x^{ls}(t) \\ \dots \\ x^{hf}(t) \\ x^{lf}(t) \end{bmatrix},$$

where, $x^s(t)$ $(n^s \times 1)$ contains $n^{hs}+n^{ls}=n^s$ stock variables and $x^f(t)$ $(n^f \times 1)$ contains $n^{hf}+n^{lf}=n^f$ flow variables. For now, we assume only two frequencies are available and use m to denote the number of high-frequency observations between low frequency observations. The extension to multiple frequencies is discussed briefly later. The following table summarises the nature and dimension of the observable data, their relationship to $x(t)$ and their availability.

Observed vector		Observation	Dimension	Availability
x_t^{hs}	$=$	$x^{hs}(t)$	n^{hs}	$t=1,2,\dots,T.$
x_t^{hf}	$=$	$\int_{t-1}^{t} x^{hf}(r)dr$	n^{hf}	$t=1,2,\dots,T.$
x_t^{ls}	$=$	$x^{ls}(t)$	n^{ls}	$t=m,2m,\dots,m\lfloor T/m\rfloor.$
x_t^{lf}	$=$	$\int_{t-m}^{t} x^{lf}(r)dr$	n^{lf}	$t=m,2m,\dots,m\lfloor T/m\rfloor.$

We use $n^h=n^{hs}+n^{hf}$, $(n^l=n^{ls}+n^{lf})$, to denote the number of high (low) frequency variables in $x(t)$.

Our aim is to translate the process in $x(t)$, $t>0$ into one for the $n^{(m)}\equiv mn^h+n^l$ vector, $x_t^{(m)}$, for $t=m,2m,\dots,Tm$, where

$$x_t^{(m)\prime}=[x_t^{(m)h\prime},\ x_t^{ls\prime},\ x_t^{lf\prime}],$$

which contains the current and $m-1$ lags of the observed high frequency stocks and flows

$$x_t^{(m)h}=[x_t^{h\prime},\ x_{t-1}^{h\prime},\dots,\ x_{t-m+1}^{h\prime}]',$$

with $x_t^h=\left[x_t^{hs\prime},x_t^{hf\prime}\right]'$, above the observed low frequency stocks and the observed low frequency flows. The strategy follows three steps: a) produce a state space model containing the skip-sampled and (at least the high frequency) time aggregated variables together at a common high frequency; b) implement a second round of aggregation to account for the absence of observations of the low frequency variables; and c) recover a linear process describing the laws of motion for the observable data vector, a so-called exact discrete representation for the mixed frequency data. The first step requires the augmentation of the system with further variables to aggregate the flows but these will be the only increase in unobservable elements. Otherwise our approach is distinguished by only appending observables to the system, keeping it of relatively low dimension. At each stage it will be necessary to partition the state vector in different ways to highlight the evolution of the system: in a) this is between the flow aggregators and the rest; in b) the important distinction is between high and low frequency; while in c) it is between observable and unobservable state variables.

2.2 Continuous to high frequency discrete time

As in Thornton and Chambers (2017), we append the state vector with n^f extra elements, partitioned without loss of generality $y_0(t) = [y_0^{h\prime}(t), y_0^{l\prime}(t)]'$, to capture the aggregated flows, using the relationship

$$Dy_0(t) = x^f(t) \Rightarrow y_0(t) - y_0(t-j) = \int_{t-j}^{t} x^f(s)ds. \tag{3}$$

It is most convenient to insert $y_0^h(t)$ ($y_0^l(t)$) immediately beneath $x^{hs}(t)$ ($x^{ls}(t)$), thereby rewriting the $np + n^f$ state vector as

$$\tilde{y}(t) = [x^{hs\prime}(t), y_0^{h\prime}(t), x^{ls\prime}(t), y_0^{l\prime}(t), x^f(t)', y_2(t)', \ldots, y_p(t)']'.$$

The consequence of (3) is to interject n^{hf} (and n^{lf}) null columns into A and n^{hf} (and n^{lf}) rows into (2) all of which are null apart from a single 1 in each row, picking out $x^{hf}(t)$ (or $x^{lf(t)}$), some $n^f + n^{ls}$ (n^f) cells after the principal diagonal. The state equation is now

$$D\tilde{y}(t) = \tilde{a} + \tilde{A}\tilde{y}(t) + \tilde{V}e(t), \tag{4}$$

see the appendix for further details. The solution to (4), conditional on $\tilde{y}(0)$, is

$$\tilde{y}(t) = e^{\tilde{A}t}\tilde{y}(0) + \int_0^t e^{\tilde{A}(t-s)} \left[\tilde{a} + \tilde{V}e(s)\right] ds, \quad t > 0, \tag{5}$$

where $e^A = I + A + A^2/2! + A^3/3! + \ldots$ and it follows that

$$\tilde{y}(t) = \tilde{c} + \tilde{C}\tilde{y}(t-1) + u(t), \quad u(t) = \int_{t-1}^{t} \tilde{C}(t-s)\tilde{V}e(s)ds, \tag{6}$$

for $t = 1, \ldots, T$, where $\tilde{C}(r) = e^{r\tilde{A}}$, $\tilde{C} = \tilde{C}(1)$, $\tilde{c} = \left[\int_0^1 \tilde{C}(r)dr\right]\tilde{a}$. Using Lemma A1 and the definition of the matrix exponential it is easy to show that columns $n^{hs} + 1 : n^h$ of $\tilde{C}$ are null apart from ones on the principal diagonal, confirming that nothing in the system depends on $y_0^h(t-1)$ other than $y_0^h(t)$. Removing the ones from these columns is equivalent to subtracting $y_0^h(t-1)$ from $y_0^h(t)$, thereby, from (3), introducing the high frequency flows into a new state vector

$$\tilde{\xi}_t^{(1)} = \left[x^{hs\prime}(t),\ \int_{t-1}^{t} x^{hf}(r)'dr,\ x^{ls}(t)',\ y_0^{lf\prime}(t),\ x^{f\prime}(t),\ y_2(t)', \ldots,\ y_p(t)'\right]',$$

with state equation

$$\tilde{\xi}_t^{(1)} = \tilde{c} + \tilde{\Phi}\tilde{\xi}_{t-1}^{(1)} + u(t). \tag{7}$$

The simultaneous redefinition of $\tilde{\xi}_{t-1}^{(1)}$ on the right hand side of (7) is permitted by the corresponding null columns in $\tilde{\Phi}$, indicating that x_{t-1}^{hf} has no influence on the system.

At this point $\tilde{\xi}_t^{(1)}$ contains x_t^{hs}, x_t^{hf} and x_t^{ls}. The low frequency stocks are yet to be aggregated from $y_0^{lf}(t)$ but it is worth noting that columns $n^h + n^{ls} + 1 : n$ are also null apart from ones on the principal diagonal.

2.3 High to mixed frequency

In this part of proceedings, we draw the distinction between the high and low frequency elements of $\tilde{\xi}_t^{(1)}$. The move to mixed frequency is then a specialisation [2] of the method outlined for discrete time linear processes in Thornton (2019). It will translate (7) into a system with a state vector

$$\xi_t^{(m)} = [\ x_t^{(m)\prime}, \quad y_t^{(m)\prime}\ ]', \tag{8}$$

with $y_t^{(m)} = \left[x^{f\prime}(t), y_2(t)', \ldots, y_p(t)'\right]'$ the unobservable elements of the state vector which will remain of dimension $b \equiv (p-1)n + n^f$ regardless of the magnitude of m.

The method for transforming these relationships to a mixed frequency process comes simply from splitting equation (7) into its high frequency observables, x_t^h, from the low frequency observables and the unobservable variables, $\tilde{\xi}_t^l = \left[x^{ls\prime}(t), y_0^{lf\prime}(t), y_t^{(m)\prime}\right]'$,

$$\begin{aligned} x_t^h &= \tilde{c}_h + \tilde{\Phi}_{hh} x_{t-1}^h + \tilde{\Phi}_{hl} \tilde{\xi}_{t-1}^l + u_t^h, & (9) \\ \tilde{\xi}_t^l &= \tilde{c}_l + \tilde{\Phi}_{lh} x_{t-1}^h + \tilde{\Phi}_{ll} \tilde{\xi}_{t-1}^l + u_t^l, & (10) \end{aligned}$$

where we have similarly partitioned

$$\tilde{\Phi} = \begin{bmatrix} \tilde{\Phi}_{hh} & \tilde{\Phi}_{hl} \\ \tilde{\Phi}_{lh} & \tilde{\Phi}_{ll} \end{bmatrix}, \ \tilde{c} = \begin{bmatrix} \tilde{c}_h \\ \tilde{c}_l \end{bmatrix}, \ u_t = \begin{bmatrix} u_t^h \\ u_t^l \end{bmatrix} \equiv \begin{bmatrix} \Theta^h \\ \Theta^l \end{bmatrix} u_t,$$

and Θ^h (Θ^l) contains the top n^h (bottom $n(p-1) + n^f + n^l$) rows of the $np + n^f$ identity matrix. Lagging (10) and substituting into (9) and (10) a total of $m-1$ times then produces (see Thornton (2019) for discussion of the details) the mixed frequency system

$$\tilde{\xi}_t^{(m)} = \tilde{c}^{(m)} + \tilde{\Phi}^{(m)} \tilde{\xi}_{t-1}^{(m)} + u_t^{(m)}, \tag{11}$$

where $\tilde{\xi}_{t-1}^{(m)} \equiv [\ x_{t-1}^{(m)h\prime} \quad \tilde{\xi}_{t-m}^{l\prime}\ ]'$ is constructed with an unusual pattern of lagging, $u_t^{(m)} = [u_t', u_{t-1}', \ldots, u_{t-m+1}']'$ and

$$\tilde{\Phi}^{(m)} = \begin{bmatrix} \tilde{\Phi}_{hh} & \tilde{\Phi}_{hl}\tilde{\Phi}_{lh} & \tilde{\Phi}_{hl}\tilde{\Phi}_{ll}\tilde{\Phi}_{lh} & \ldots & \tilde{\Phi}_{hl}\tilde{\Phi}_{ll}^{m-2}\tilde{\Phi}_{lh} & \tilde{\Phi}_{hl}\tilde{\Phi}_{ll}^{m-1} \\ 0 & \tilde{\Phi}_{hh} & \tilde{\Phi}_{hl}\tilde{\Phi}_{lh} & \ldots & \tilde{\Phi}_{hl}\tilde{\Phi}_{ll}^{m-3}\tilde{\Phi}_{lh} & \tilde{\Phi}_{hl}\tilde{\Phi}_{ll}^{m-2} \\ 0 & 0 & \tilde{\Phi}_{hh} & \ldots & \tilde{\Phi}_{hl}\tilde{\Phi}_{ll}^{m-4}\tilde{\Phi}_{lh} & \tilde{\Phi}_{hl}\tilde{\Phi}_{ll}^{m-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & \tilde{\Phi}_{hh} & \tilde{\Phi}_{hl} \\ \tilde{\Phi}_{lh} & \tilde{\Phi}_{ll}\tilde{\Phi}_{lh} & \tilde{\Phi}_{ll}^2\tilde{\Phi}_{lh} & \ldots & \tilde{\Phi}_{ll}^{m-1}\tilde{\Phi}_{lh} & \tilde{\Phi}_{ll}^m \end{bmatrix},$$

[2] The restriction implied by the continuous time model that $p > q$ is very helpful in simplifying the general problem.

$$\tilde{c}^{(m)} = \begin{bmatrix} \tilde{c}_h + \tilde{\Phi}_{hl} \sum_{j=0}^{m-2} \tilde{\Phi}_{ll}^j \tilde{c}_l \\ \tilde{c}_h + \tilde{\Phi}_{hl} \sum_{j=0}^{m-3} \tilde{\Phi}_{ll}^j \tilde{c}_l \\ \vdots \\ \tilde{c}_h + \tilde{\Phi}_{hl} \tilde{c}_l \\ \tilde{c}_h \\ \sum_{j=0}^{m-1} \tilde{\Phi}_{ll}^j \tilde{c}_l \end{bmatrix},$$

and

$$\Theta^{(m)} = \begin{bmatrix} \Theta_h & \tilde{\Phi}_{hl}\tilde{\Theta}_l & \tilde{\Phi}_{hl}\tilde{\Phi}_{ll}\tilde{\Theta}_l & \dots & \tilde{\Phi}_{hl}\tilde{\Phi}_{ll}^{m-2}\tilde{\Theta}_l \\ 0 & \Theta_h & \tilde{\Phi}_{hl}\tilde{\Theta}_l & \dots & \tilde{\Phi}_{hl}\tilde{\Phi}_{ll}^{m-3}\tilde{\Theta}_l \\ 0 & 0 & \Theta_h & \dots & \tilde{\Phi}_{hl}\tilde{\Phi}_{ll}^{m-4}\tilde{\Theta}_l \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \Theta_h \\ \tilde{\Theta}_l & \tilde{\Phi}_{ll}\tilde{\Theta}_l & \tilde{\Phi}_{ll}^2\tilde{\Theta}_l & \dots & \tilde{\Phi}_{ll}^{m-1}\tilde{\Theta}_l \end{bmatrix}.$$

We pause to remark upon a similarity between equations (11) and (7) which is such that, if multiple frequencies were required, the intervening steps could be repeated to perform additional aggregations as required, see Thornton (2019).

All that remains is to recover the low frequency flow variables from $\tilde{\xi}_t^l$ and $\tilde{\xi}_{t-m}^l$. It is not too difficult to verify that since the columns of $n^h + n^{ls} + 1 : n$ of $\tilde{\Phi}$, relating to $y_0^{lf}(t-1)$, are null apart from ones on the principal diagonal, so are columns $mn^h + n^{ls} + 1 : n^{(m)}$ of $\tilde{\Phi}^{(m)}$: since columns $n^{ls} + 1 : n^l$ of $\tilde{\Phi}_{ll}$ are null apart from ones on the principal diagonal so are those of $\tilde{\Phi}_{ll}^j$ by Lemma A2; and, as those columns of $\tilde{\Phi}_{hl}$ are null so are those of $\tilde{\Phi}_{hl}\tilde{\Phi}_{ll}^j$ by Lemma A3. This is to say that the coefficients on $y_{0,t-m}^{lf}$ in $\tilde{\Phi}^{(m)}$ are null apart from the block relating to $y_{0,t}^{lf}$, which contains the identity matrix. In an identical move to that performed on $y_{0,t-1}^{hf}$, removing that identity matrix is equivalent to taking $y_{0,t-m}^{lf}$ over to the left hand side and thereby introducing the variable $y_{0,t}^{lf} - y_{0,t-m}^{lf} = x_t^{(m)lf}$ into the state vector, $\xi_t^{(m)} = [\begin{array}{cc} x_t^{(m)h\prime} & \xi_t^{l\prime} \end{array}]'$ defined by equation (8). The resulting system

$$\xi_t^{(m)} = \tilde{c}^{(m)} + \Phi^{(m)}\xi_{t-1}^{(m)} + \Theta^{(m)}u_t^{(m)} \tag{12}$$

where $\Phi^{(m)}$ differs from $\tilde{\Phi}^{(m)}$ only in that the aforementioned columns, now relating to x_{t-m}^{lf}, are null. This feature would make it impossible to reconstruct $y_{0,t}$ from the other elements in the state vector, were it not observable. Methods to reconstruct the other elements of $\xi_t^{(m)}$ are the focus of the next section.

2.4 ARMA reconstruction

For this final part of proceedings the most important distinction is between the observable and the unobservable elements of $\xi_t^{(m)}$, since the exact discrete representation for the linear system (12) results from using lags of the former and disturbance to replace the latter. The observable data, $x_t^{(m)} = S_1\xi_t^{(m)}$, $t = m, 2m, \dots, \lfloor T/m \rfloor m$, where $S_1 = [I, 0]$ has $n^{(m)}$ rows.

The unobservable data, $y_t^{(m)} = S_2\xi_t^{(m)}$, where $S_2 = [0,\, I]$ has $b = (p-1)n + n^f$ rows. The task is to replace $y_t^{(m)}$ with lags of $x_t^{(m)}$, the intercept and the disturbance, noting that the high frequency observables on the right hand side of equation (12) are lagged one, rather than m periods. In order to split equation (12) into $x_t^{(m)}$ and $y_t^{(m)}$ and their lags at frequency m, we need the selection matrices S_{x^l,ξ^l} and S_{y,ξ^l} to separate the low frequency elements of the state vector, ξ_t^l, into observables and unobservables, respectively. Using these to collect terms to write

$$N_a x_t^{(m)} = N_b x_{t-m}^{(m)} + \Phi_{12} y_{t-m}^{(m)} + \tilde{c}_1^{(m)} + \Theta_1^{(m)} u_t^{(m)}, \tag{13}$$

$$y_t^{(m)} = N_c x_t^{(m)} + N_d x_{t-m}^{(m)} + \Phi_{22} y_{t-m}^{(m)} + \tilde{c}_2^{(m)} + \Theta_2^{(m)} u_t^{(m)}, \tag{14}$$

where $\tilde{c}_i^{(m)} = S_i\tilde{C}^{(m)}$, $\Theta_i^{(m)} = S_i\Theta^{(m)}$, $\Phi_{ij} = S_i\Phi^{(m)}S_j'$,

$$N_a = -\begin{bmatrix} -I_{n^h} & \tilde{\Phi}_{hh} & \tilde{\Phi}_{hl}\tilde{\Phi}_{lh} & \tilde{\Phi}_{hl}\tilde{\Phi}_{ll}\tilde{\Phi}_{lh} & \dots & \tilde{\Phi}_{hl}\tilde{\Phi}_{ll}^{m-3}\tilde{\Phi}_{lh} & 0 \\ 0 & -I_{n^h} & \tilde{\Phi}_{hh} & \tilde{\Phi}_{hl}\tilde{\Phi}_{lh} & \dots & \tilde{\Phi}_{hl}\tilde{\Phi}_{ll}^{m-4}\tilde{\Phi}_{lh} & 0 \\ 0 & 0 & -I_{n^h} & \tilde{\Phi}_{hh} & \dots & \tilde{\Phi}_{hl}\tilde{\Phi}_{ll}^{m-5}\tilde{\Phi}_{lh} & 0 \\ \vdots & \vdots & \ddots & \ddots & & & \vdots \\ 0 & 0 & 0 & 0 & \dots & -I_{n^h} & 0 \\ 0 & S_{x,\xi}\tilde{\Phi}_{lh} & S_{x,\xi}\tilde{\Phi}_{ll}\tilde{\Phi}_{lh} & S_{x,\xi}\tilde{\Phi}_{ll}^2\tilde{\Phi}_{lh} & \dots & S_{x,\xi}\tilde{\Phi}_{ll}^{m-2}\tilde{\Phi}_{lh} & -I_{n^l} \end{bmatrix},$$

$$N_b = \begin{bmatrix} \tilde{\Phi}_{hl}\tilde{\Phi}_{ll}^{m-2}\tilde{\Phi}_{lh} & 0 & \tilde{\Phi}_{hl}\tilde{\Phi}_{ll}^{m-1}S_{x,\xi}' \\ \tilde{\Phi}_{hl}\tilde{\Phi}_{ll}^{m-3}\tilde{\Phi}_{lh} & 0 & \tilde{\Phi}_{hl}\tilde{\Phi}_{ll}^{m-2}S_{x,\xi}' \\ \tilde{\Phi}_{hl}\tilde{\Phi}_{ll}^{m-4}\tilde{\Phi}_{lh} & 0 & \tilde{\Phi}_{hl}\tilde{\Phi}_{ll}^{m-3}S_{x,\xi}' \\ \vdots & \vdots & \vdots \\ \tilde{\Phi}_{hh} & 0 & \tilde{\Phi}_{hl}S_{x,\xi}' \\ S_{x,\xi}\tilde{\Phi}_{ll}^{m-1}\tilde{\Phi}_{lh} & 0 & S_{x,\xi}\tilde{\Phi}_{ll}^{m}S_{x,\xi}' \end{bmatrix},$$

$$N_c = S_{y,\xi}\begin{bmatrix} 0 & \tilde{\Phi}_{lh} & \tilde{\Phi}_{ll}\tilde{\Phi}_{lh} & \tilde{\Phi}_{ll}^2\tilde{\Phi}_{lh} & \dots & \tilde{\Phi}_{ll}^{m-2}\tilde{\Phi}_{lh} & 0 \end{bmatrix},$$

and

$$N_d = S_{y,\xi}\begin{bmatrix} \tilde{\Phi}_{ll}^{m-1}\tilde{\Phi}_{lh} & 0 & \tilde{\Phi}_{ll}^{m}S_{x,\xi}' \end{bmatrix}.$$

In order to produce an expression for $y_{t-m}^{(m)}$ that may be substituted into (13) define the following vectors of lagged observable variables, for g lags,

$$\bar{x}_t = [x_{t-m}^{(m)\prime},\, \dots,\, x_{t-(g+1)m}^{(m)\prime}]';$$

unobservables,

$$\bar{y}_t = [y_{t-m}^{(m)\prime},\, \dots,\, y_{t-(g+1)m}^{(m)\prime}]';$$

and disturbances,

$$\bar{u}_t = [u_{t-m}^{(m)\prime},\, \dots,\, u_{t-gm}^{(m)\prime}]'.$$

The resulting system of $g(n^{(m)}+b)$ equations in $(g+1)b$ unknowns is

$$\bar{M}\bar{y}_t = \bar{N}\bar{x}_t + \bar{H} + \bar{\Theta}\bar{u}_t, \tag{15}$$

where $\bar{H} = [i_g' \otimes \tilde{c}_1^{(m)\prime}, i_g' \otimes \tilde{c}_2^{(m)\prime}]'$, with i_g denoting a g vector of ones,

$$\begin{aligned}\bar{N} &= \begin{bmatrix} N_1 & N_2 & N_3 & \dots & N_g & N_{g+1} \end{bmatrix} \\ &= \left[\begin{array}{cccccc} -N_a & N_b & 0 & \dots & 0 & 0 \\ 0 & -N_a & N_b & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -N_a & N_b \\ \hline N_c & N_d & 0 & \dots & 0 & 0 \\ 0 & N_c & N_d & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \dots & N_c & N_d \end{array}\right].\end{aligned}$$

$$\begin{aligned}\bar{\Theta} &= \begin{bmatrix} \bar{\Theta}_1 & \bar{\Theta}_2 & \bar{\Theta}_3 & \dots & \bar{\Theta}_{g-1} & \bar{\Theta}_g \end{bmatrix} \\ &= \left[\begin{array}{cccccc} \Theta_1^{(m)} & 0 & 0 & \dots & 0 & 0 \\ 0 & \Theta_1^{(m)} & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & \Theta_1^{(m)} \\ \hline \Theta_2^{(m)} & 0 & 0 & \dots & 0 & 0 \\ 0 & \Theta_2^{(m)} & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & \Theta_2^{(m)} \end{array}\right].\end{aligned}$$

Equation (15) can only be solved for $\bar{y}$ if the $g(n^{(m)}+b) \times (g+1)b$ matrix

$$\bar{M} = \left[\begin{array}{cccccc} 0 & -\Phi_{12} & 0 & \dots & 0 & 0 \\ 0 & 0 & -\Phi_{12} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & -\Phi_{12} \\ \hline I_b & -\Phi_{22} & 0 & \dots & 0 & 0 \\ 0 & I_b & -\Phi_{22} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \dots & I_b & -\Phi_{22} \end{array}\right],$$

has full column rank, a necessary condition for which is that $g \geq b/n^{(m)}$. Provided that these conditions are met, the top b rows of $\bar{y}_t$, $y_{t-m}^{(m)}$, may be substituted into (13), thereby giving an expression for $x_t^{(m)}$ in terms of its own lags at frequency m and lags of the disturbance term u_t. That expression is summarised in the following Theorem.

Theorem 1 Provided that $\bar{M}$ has full column rank then for matrix R such that $R\bar{M}$ is non-singular the observed vector $x^{(m)}$ then satisfies the discrete time ARMAX$(g+1,g)$ system

$$N_a x_t^{(m)} = F_1 x_{t-m}^{(m)} + \ldots + F_{g+1} x_{t-(g+1)m} G_0 + \eta_t, \quad t = (g+2)m, \ldots, \lfloor T/m \rfloor m,$$

where $F_1 = N_b + \Phi_{12}\hat{M}N_1$ then $F_j = \Phi_{12}\hat{M}N_j$ for $(j = 2, \ldots, g+1)$; and, $G_0 = \tilde{c}_1^{(m)} + \Phi_{12}\hat{M}\bar{H}$,

$$\begin{aligned} \hat{M} &= [I_b, 0_{b\times gb}][R\bar{M}]^{-1}R = [\hat{M}_1, \hat{M}_2], \\ \hat{M}_1 &= [\hat{M}_{1,1}, \hat{M}_{1,2}, \ldots, \hat{M}_{1,g}], \\ \hat{M}_2 &= [\hat{M}_{2,1}, \hat{M}_{2,2}, \ldots, \hat{M}_{2,g}], \end{aligned}$$

the matrices $\hat{M}_{1,i}$ and $\hat{M}_{2,i}$ $(i = 1, \ldots, m)$ being $b \times n^{(m)}$ and $b \times b$, respectively.

$$\Gamma_j^{(m)} = E(\eta_t \eta_{t-jm}') = \begin{cases} \sum_{i=j}^{g} C_i \Omega_u C_{i-j}', & j = 0, \ldots, g, \\ 0, & j > g, \end{cases}$$

where $C_0 = S_1$, $C_j = \Phi_{12}(\hat{M}_{1,j}S_1 + \hat{M}_{2,j+1}S_2)$ $(j = 1, \ldots, g)$, and

$$\Omega_u = \Theta^{(m)} \left[\int_{t-1}^{t} \tilde{C}(t-r)\tilde{V}\Sigma\tilde{V}'\tilde{C}(t-r)'dr \right] \Theta^{(m)\prime}.$$

Proof The proof follows that of Theorem 1 in Chambers (1999), allowing the possibility that $\bar{M}$ may not be square. Once $\bar{M}$ has full column rank there exists a matrix R such that $R\bar{M}$ is non-singular. We may then get y_{t-m} from the first n^{ξ} equations of

$$\bar{y}_t = [R\bar{M}]^{-1}R[\bar{N}\bar{x}_t + \bar{H} + \bar{\Theta}\bar{u}_t].$$

Substituting back into equation (13) gives the expression in the Theorem. The covariance structure of the disturbance is due to the white noise properties of the continuous time disturbance $e(t)$. $\square$

Theorem 1 generalises the expressions in Chambers (2016) to higher order continuous time ARMA (p,q) models. The expressions for N_a and N_b are familiar but with a higher order model there are also unobservables in our state representation for which we must later solve. The same is true of mixed stock-flow processes, which we chose to handle in a similar manner but Chambers solves based on an assumption that a sub-matrix of A is non-singular. This type of assumption is relatively common in handling mixed stock flow data and corresponds to assumptions on the rank of $\bar{M}$. It should be noted that the quasi-upper triangular form of N_a, resulting from the form of $\Phi^{(m)}$, with the identity matrix on the principal diagonal provides a convenient method to express the high frequency variables in terms of their own lags at high-frequency in between observations of the low frequency variables. This has the effect of providing expressions valid at all points in the observation cycle, not simply every m periods, overcoming the so-called ‘ragged edge’ problem.

The disturbance vector on this mixed frequency process, η_t, has a moving average representation of order g. Its covariance structure features both the moving average parameters of the continuous time model, via $\tilde{V}$, and the autoregressive parameters, via both the matrices $\tilde{C}$ and C_j. This transforms the moving average structure of η from being a nuisance to an important source of information about the autoregressive parameters to be utilised in estimation.

The form of the representation will depend on the choice of matrix R. A natural choice is to set $R = [\bar{M}'\bar{M}]^{-1}\bar{M}'$, the Moore-Penrose inverse, as was the case in the following sections, but others may be available and may lead to lower order representations. For example, if it is possible to chose an R such that $RN_{g+1} = 0$ then the representation would only be based on g low frequency lags. Given that all but n columns of both N_b and N_d are null, N_{g+1} has a nullspace of dimension at least $g(n^m + b) - n$ from which future research may determine it possible to construct the columns of a suitable R.

3 Estimation techniques

Gaussian estimation, see Bergstrom (1983, 1990), uses the exact observed ARMA representation derived in Theorem 1 as the foundation of a quasi-maximum likelihood estimator. It involves calculations of the $n^{(m)}$ residual vector, η_t, and its implied covariance structure, given a particular point in the parameter space. With a sample of size T, there will be $\lfloor T/m \rfloor - g - 1$ completed sample cycles and a ragged edge of size $T - m\lfloor T/m \rfloor$ for which only high frequency observations are available. We define the vector of discrete time disturbances, $\eta = (\eta'_{(g+2)m}, \eta'_{(g+3)m} \dots, \eta'_{T^*})'$, where $T^* = m\lceil T/m \rceil$. In the event that T is not an integer multiple of m the bottom $n^l + m(\lceil T/m \rceil - T)n^h$ rows of η_{T^*} are left null so that η has $n^* = Tn^h + \lfloor T/m \rfloor n^l - (g+1)n^{(m)}$ non-zero elements.

The covariance matrix, $\Omega_\eta = E(\eta\eta')$, has a block Toeplitz structure with ij'th block denoted by the n matrix

$$\Omega_{\eta,ij} = \begin{cases} \Gamma^{(m)}_{i-j}, & |i-j| \leq g, \\ 0, & |i-j| > g, \end{cases}$$

noting that $\Gamma^{(m)}_{-j} = \Gamma^{(m)\prime}_j$ with $\Gamma^{(m)}_j$ defined in Theorem 1. A quasi-maximum likelihood estimator may be obtained by imagining that η has a multivariate normal distribution, enabling the likelihood to be evaluated as

$$\log L = -\frac{n^*}{2}\log(2\pi) - \frac{1}{2}\log|\Omega_\eta| - \frac{1}{2}\eta'\Omega_\eta^{-1}\eta.$$

As pointed out in Bergstrom (1983, 1990), the sparse nature of Ω_η makes it possible to accelerate the calculation of this likelihood. Since Ω_η is positive definite and symmetric we can find a lower triangular matrix, U, such that

$$UU' = \Omega_\eta,$$

with the sparse nature of Ω_η reflected in the sparse nature of U, further details are given in Bergstrom (1983) and in Thornton and Chambers (2017). A vector of normalised residuals,

ζ such that $E(\zeta) = 0$ and $E(\zeta\zeta') = I_{n^*}$, satisfying $U\zeta = \eta$ may be recovered by a recursive operation and it follows straightforwardly that,

$$\log L = -\frac{n^*}{2}\log(2\pi) - \frac{1}{2}\zeta'\zeta - \log(|U|),$$

where $\log(|U|)$ is easily calculated as the sum of the terms on the principle diagonal of U. Calculation of U involves inverting a maximum of $T - g$ lower triangular matrices of dimension $n^{(m)}$, but in practice U_{ij} and $U_{i+1,j+1}$ often converge quickly, removing the need to calculate the rows of U further, see Bergstrom (1990, ch 7). The non-zero blocks in block row i contain coefficient matrices for a moving average representation of η_{im}. In addition to its use in computing the log-likelihood function, the vector of normalised residuals $\zeta = [\zeta_1', \zeta_2', \ldots, \zeta_T']'$ can be used to conduct a general test of dynamic specification. Bergstrom (1990, chapter 7) proposed a portmanteau-type test statistic of the form

$$S_l = \frac{1}{n(T-l)}\sum_{r=1}^{l}\left(\sum_{t=l+1}^{T}\zeta_t'\zeta_{t-r}\right)^2,$$

which, under the null hypothesis that the model is correctly specified, has an approximate χ_l^2 distribution for sufficiently large l and $T - l$, where l ($> p$) denotes the number of lags used.

4 Simulation

To show the effectiveness of these techniques we test them against simulated data. We simulated 10,000 replications of 240 (high frequency) observation points for a two variable system containing a stock variable, observed every period, and a flow aggregated over every three - mirroring 20 years of monthly data in a stock and quarterly data for a flow. Estimates were then made of the continuous time parameters using the above methods and compared to those if only quarterly data were used, that is to say with two out of every three stock observations discarded. By way of benchmarking, the performance of these estimators is reported relative to an infeasible high frequency estimator, where the analyst is granted access to all data at the monthly frequency.

Each experiment is based on a random set of autoregressive and moving average parameters, subject to the constraints of miniphaseness and stationarity. We allowed for contemporaneous correlation in the continuous time disturbance vector, setting the covariance matrix to

$$\Sigma = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix},$$

in all experiments. The reported figures are for estimates of the Choleski decomposition

$$Q = \begin{bmatrix} s_{11} & 0 \\ s_{21} & s_{22} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0.5 & 0.86603 \end{bmatrix},$$

with $\Sigma = QQ'$

Table 1 contains the results for the CAR(1), which corresponds to (1) with $p = 1$. Data were simulated using parameters

$$A_0 = \begin{bmatrix} -0.6427 & -0.1004 \\ -0.9534 & -0.5819 \end{bmatrix},$$

which has eigenvalues [3] -0.9232 and -0.3014. It is not surprising that both the bias and root

Table 1: Bias and RMSE for estimates of the continuous time ARMA (1, 0) model relative to the infeasible high frequency estimator

	bias		RMSE	
frequency	low	mixed	low	mixed
A^{hh}	3.0866	0.1659	17.2331	1.0716
A^{hl}	1.6084	6.3118	3.3678	1.2646
A^{lh}	15.1196	-4.2985	31.9687	1.5013
A^{ll}	43.9308	-21.8481	6.3845	1.4950
s_{11}	1.0694	0.8939	9.7174	0.9643
s_{21}	1.1539	0.7250	6.5325	0.8688
s_{22}	1.1329	1.3177	1.4699	1.3178
mean	9.5859	-2.3903	10.9534	1.2119

mean square error are typically larger in both than in the infeasible model, particularly for parameters A_{lh} and A_{ll} related to the low frequency series. The mixed frequency model does, however, perform far closer to the infeasible estimator than the the low frequency estimator does, with a root mean square error typically less than 50 per cent larger than the infeasible estimator, compared to the large multiples reported by the low frequency estimator. This difference is at its most pronounced for parameters that measure the effect of the high frequency variables within the system, A_{lh} and A_{hh}. Table 1 confirms and extends the results in Chambers (2016), who explored the continuous time AR(1) with pure stock variables, that the performance of the mixed frequency estimator often lies closer to that of the infeasible estimator than to that of the low frequency estimator.

Table 2 covers the stationary CARMA (2,1) model in which

$$A_0 = \begin{bmatrix} -0.6671 & -0.6793 \\ -0.1466 & -0.5488 \end{bmatrix}, \; A_1 = \begin{bmatrix} -0.2448 & -0.1167 \\ -0.2328 & -0.5757 \end{bmatrix},$$

with eigenvalues $-0.4705 \pm 0.6603i$ and $-0.1348 \pm 0.3936i$. Although the performance of the mixed frequency estimator is adversely affected, it remains preferable to that of the low frequency estimator in terms of root mean square error.

[3]Eigenvalues with negative real parts correspond to a stationary discrete time model with roots outside the unit circle.

Table 2: Bias and RMSE for estimates of the continuous time ARMA (2, 0) model relative to the infeasible high frequency estimator

	bias		RMSE	
frequency	low	mixed	low	mixed
A_1^{hh}	17.3675	52.7681	6.8045	5.9426
A_1^{hl}	35.7049	74.7739	10.1851	6.9697
A_1^{lh}	77.2112	67.8104	8.6129	2.8974
A_1^{ll}	47.1320	73.9560	12.1618	7.5928
A_2^{hh}	19.9365	31.6971	23.1845	6.1825
A_2^{hl}	20.2668	10.1674	12.7794	2.3802
A_2^{lh}	29.2322	33.9392	22.7937	7.1593
A_2^{ll}	24.3162	42.6729	20.5259	12.6005
s_{11}	6.4019	9.5575	13.6134	7.8108
s_{21}	2.9950	2.4619	6.6859	2.6348
s_{22}	0.9920	−3.3096	3.4317	5.2601
mean	25.5960	36.0450	12.7981	6.1301

5 Application:a cointegrated model of stock prices and dividends

To illustrate the applicability of these techniques, we now consider a multivariate model of stock prices and dividends using the popular dataset of Shiller (2000)[4], which provides a time series of monthly figures for stock prices and dividends for the Standard and Poor's index. The monthly figures for dividends are, however, based on the linear interpolation of quarterly figures. The above framework allows the estimation of a model of monthly stock and quarterly dividends, with the observed dividend payout, $d_t = \int_{t-1}^{t} d(\tau)d\tau$, regarded as a flow variable, reflecting an observed accrual of profits over the observation period. It reprises the relationship estimated in Thornton and Chambers (2016) with the enhancement that the differing data frequencies are handled internally within the aggregation of the underlying model rather externally via interpolation.

We follow the seminal work by Campbell and Shiller (1987) in analysing the relationship between the logarithm of the stock price and the logarithm of dividends using the data spanning the period 1871–1986. The sample is chosen to pre-date the fashion for companies to remunerate investors by re-purchasing shares, thereby raising their price, rather than externally paying dividends, which is liable to disrupt the relationship. Both series display unit-root type behaviour. Following the work of Campbell and Shiller (1987) on the so-called 'present value model' it is widely postulated that, since stock prices should represent the discounted flow of future dividends, which are themselves highly persistent, the long run relationship between the two series is a form of cointegration, with the discount factor determining the cointegrating vector, see Thornton and Chambers (2016) for further discussion on the foundations of this and other models as well as the treatment of dividends.

We consider three models based on equation (1) with $x(t) = (s(t), d(t))'$. In a two

[4]The data are available at http://www.econ.yale.edu/˜shiller/data.htm.

variable continuous time system, cointegration implies that we may write, without loss of generality, $A_0 = \alpha\beta'$, where $\alpha' = [\alpha_1, \alpha_2]$ and $\beta' = [1, \beta_1]$ is a cointegrating vector such that $\beta' x(t)$ is stationary. We would expect β_1 to be negative and slightly above one in magnitude to reflect the discounting of future dividends, while α_1 and α_2 may be interpreted as speed-of-adjustment parameters, with error correction implying $\alpha_1 < 0$ and $\alpha_2 > 0$.

Estimates for the CARMA$(1,0)$ model,

$$\begin{aligned} Ds(t) &= a_{0,1} + \alpha_1 s(t) + \alpha_1\beta_1 d(t) + u_1(t), \\ Dd(t) &= a_{0,2} + \alpha_2 s(t) + \alpha_2\beta_1 d(t) + u_2(t), \end{aligned}$$

where $u(t) = [u_1(t), u_2(t)]' \sim N(0, \Sigma_u)$ and $\Sigma_u = QQ'$ with Q a lower triangular matrix, are reported in Table 3.

Table 3: Estimates of cointegrated CARMA$(1,0)$ model for stock prices and dividends

	$Ds(t)$	$Dd(t)$
a_0'	0.0000	−0.0277
	(0.0047)	(0.0064)
α'	0.0006	0.0199
	(0.0047)	(0.0032)
β'	1.0000	−1.4542
	(-)	(0.1049)
Q'	0.0420	−0.0018
	(0.0008)	(0.0012)
	0.0000	−0.0278
	(-)	(0.0009)
log L		4116.5509
S_{12} S_{20}	[0.0001]	[0.0032]
(standard errors in parentheses)		
[p-values in braces]		

The estimate of β_1 is close to -1.45, but α_1 is not statistically significantly different from zero and has the wrong sign, placing the burden of error correction within the system on dividends. Indeed the left hand column suggests that, in the absence of more short-run sophisticated dynamics, stock-prices follow a random walk. The Bergstrom S statistic is in the extreme right tail of its asymptotic distribution for both 12 and 20 lags, suggesting a higher order dynamic structure is needed.

We also report estimates of CARMA$(2,0)$ and CARMA$(2,1)$ systems in Tables 4 and 5 respectively; the latter is given by

$$D^2 s(t) = a_{0,1} + A_{1,11} Ds(t) + A_{1,12} Dd(t) + \alpha_1 s(t) + \alpha_1\beta_1 d(t) + w_1(t), \tag{16}$$

$$D^2 d(t) = a_{0,2} + A_{1,21} Ds(t) + A_{1,22} Dd(t) + \alpha_2 s(t) + \alpha_2\beta_1 d(t) + w_2(t), \tag{17}$$

where $w_1(t) = u_1(t) + \Theta_{11} Du_1(t) + \Theta_{12} Du_2(t)$ and $w_2(t) = u_2(t) + \Theta_{21} Du_1(t) + \Theta_{22} Du_2(t)$ are defined for notational convenience. The CARMA$(2,0)$ model is obtained by setting $\Theta_{ij} = 0$ $(i, j = 1, 2)$.

The addition of higher order dynamics significantly improves the fit of the model, with a likelihood ratio test preferring the CARMA $(2,0)$ over the CARMA $(1,0)$ and the two coefficients on the principal diagonal of A_1 statistically significantly different from zero.

Table 4: Estimates of cointegrated CARMA$(2,0)$ model for stock prices and dividends

	$D^2s(t)$	$D^2d(t)$
a_0'	0.0143	0.0056
	(0.0142)	(0.0260)
A_1'	-2.4555	0.1541
	(0.2196)	(0.1497)
	-0.1711	-0.5062
	(0.1755)	(0.0825)
α'	-0.0091	0.0121
	(0.0099)	(0.0024)
β'	1.0000	-1.4699
	(-)	(0.1108)
Q'	0.1302	-0.0083
	(0.0084)	(0.0073)
	0.0000	0.0168
	(-)	(0.0018)
log L		4174.0984
S_{12} S_{20}	[0.0032]	[0.0343]
(standard errors in parentheses)		
[p-values in braces]		

The speed of adjustment parameters have the expected signs and are now both statistically significant. There is still, however, some evidence of dynamic misspecification according to the Bergstrom S statistic. The inclusion of the moving average disturbance into the continuous time model appears to satisfy the specification test, albeit marginally at 12 lags. The likelihood ratio test of the restriction that the four continuous time MA parameters are jointly zero is over 20, well into the critical region for its asymptotic χ_4^2 distribution, with the individual t-ratios suggesting that it is the parameter Θ_{22} in the equation describing the law of motion for dividends that benefits most from the inclusion. In all three specifications the estimate of β_1 remains remarkably stable at between -1.45 and -1.48.

6 Conclusions

We have derived the exact discrete representation for a vector of mixed frequency data generated by a continuous time ARMA (p, q) model, featuring both stock and flow variables at both frequencies. This extends the framework in Chambers (2016) in both the order of the model and the type of aggregation scheme. The advantage of this approach is that it enables a relatively computationally efficient evaluation of the quasi-likelihood. Simulation evidence suggest that this estimator out-performs more traditional techniques based on a concentration of the data to a single low frequency.

7 Data availability statement

The data used in the application are included as a supplementary file. They were downloaded from Robert Shiller's website, *http://www.econ.yale.edu/˜shiller/data.htm* and are discussed

Table 5: Estimates of cointegrated CARMA$(2,1)$ model for stock prices and dividends

	$D^2 s(t)$	$D^2 d(t)$
a_0'	0.0180	0.0084
	(0.0090)	(0.0123)
A_1'	-1.8157	0.1827
	(0.4798)	(0.1666)
	-0.3068	-0.0875
	(0.2297)	(0.0535)
α'	-0.0123	0.0037
	(0.0063)	(0.0014)
β'	1.0000	-1.4790
	(-)	(0.1092)
Θ_1'	-0.2884	-0.4693
	(0.6008)	(0.5330)
	-1.0927	-6.0214
	(6.3414)	(2.3683)
Q'	-0.0987	0.0085
	(0.0233)	(0.0091)
	0.0000	0.0039
	(-)	(0.0013)
log L		4185.0373
S_{12} S_{20}	[0.0517]	[0.1911]
(standard errors in parentheses)		
[p-values in braces]		

in Shiller (2000).

REFERENCES

Bergstrom, A.R. (1983) Gaussian estimation of structural parameters in higher order continuous time dynamic models. *Econometrica* 51, 117–152.

Bergstrom, A.R. (1985) The estimation of parameters in nonstationary higher-order continuous-time dynamic models. *Econometric Theory* 1, 369–385.

Bergstrom, A.R. (1990) *Continuous Time Econometric Modelling.* Oxford University Press.

Bergstrom, A.R. & K.B. Nowman (2007) *A Continuous Time Econometric Model of the United Kingdom with Stochastic Trends.* Cambridge University Press, Cambridge.

Campbell, J.Y. & R.J. Shiller (1987) Cointegration and tests of present value models. *Journal of Political Economy* 95, 1062–1088.

Chambers, M.J. (1999) Discrete time representation of stationary and non-stationary continuous time systems. *Journal of Economic Dynamics and Control* 23, 619–639.

Chambers, M.J. & M.A. Thornton (2012) Discrete time representations of continuous time ARMA processs. *Econometric Theory* 28, 219–238.

Chambers, M.J. (2016) The estimation of continuous time models with mixed frequency data. *Journal of Econometrics* 193, 390–404.

Chen, B. & P.A. Zadrozny, (1998) An extended Yule-Walker method for estimating a vector autoregressive model with mixed-frequency data. *Advances in Econometrics* 13, 47–73

Foroni, C. & M. Marcellino (2013) A survey of econometric methods for mixed-frequency data. *EUI working paper 13/02.*

Foroni, C., Marcellino, M. & C. Schumacher (2015) Unrestricted mixed data sampling (MIDAS): MIDAS regressions with unrestricted lag polynomials. *Journal of the Royal Statistal Society series A* 178, 57-82.

Ghysels, E., Santa-Clara, P. & R. Valkanov (2004) *The MIDAS touch: MIxed DAta Sampling regression models.* mimeo, Chapel Hill, N.C.

Ghysels, E. (2016) Macroeconomics and the reality of mixed-frequency data. *Journal of Econometrics* 193, 294–314.

Lütkepohl, H. (1987) *Forecasting Aggregated Vector ARMA Processes.* Springer.

Marcellino, M. (1996) Some temporal aggregation issues in empirical analysis. Economics Department Working Paper 96-39, University of California, San Diego, CA.

Marcellino, M. (1999) Some consequences of temporal aggregation in empirical analysis. *Journal of Business and Economic Statistics* 17, 129–136.

Seong, B., Ahn, S.K. & P.A. Zadrozny (2013) Estimation of vector error correction models with mixed-frequency data. (2013) *Journal of Time Series Analysis.* 34, 194-205.

Schumacher, C. (2016) A comparison of MIDAS and bridge equations. *International Journal of Forecasting* 32, 257-270.

Shiller, R.J. (2000) *Irrational Exuberance.* Princeton University Press.

Thornton, M.A. & M.J. Chambers (2016) The exact discretisation of CARMA models with applications in finance. *Journal of Empirical Finance* 38, 739-61.

Thornton, M.A. & M.J. Chambers (2017) Continuous time ARMA processes: discrete time representation and likelihood evaluation. *Journal of Economic Dynamics and Control* 79, 48-65.

Thornton, M.A. (2019) Mixed Time Aggregation of Dynamic Multivariate Linear Processes. *University of York working paper.*

Zadrozny, P.A. (1988) Gaussian likelihood of continuous-time ARMAX models when data are stocks and flows at different frequencies. *Econometric Theory* 4, 108–124.

Zadrozny, P.A. (2016) Extended Yule–Walker identification of VARMA models with single or mixed-frequency data *Journal of Econometrics* 193, 438–446.

[dataset] *http://www.econ.yale.edu/˜shiller/data.htm..*

Detail on the matrices in equation (4)

The decision to insert $y_0^h(t)$ $(y_0^l(t))$ immediately beneath $x^{hs}(t)$ $(x^{ls}(t))$, in other words in the order that x_t^{hf} (x_t^{lf}) will ultimately appear in the state matrix, simplifies the narrative at the cost of splitting up pre-defined blocks of the A matrix. In general we partition the A_k matrices in (4).

$A_k = \begin{pmatrix} A_k^{hs} & A_k^{ls} & A_k^{hf} & A_k^{lf} \end{pmatrix}, k = 0, 1, \ldots p-2,$

where A_k^j is $n \times n^j$ for $j = hs, ls, hf, lf$, while similarly partitioning A_{p-1} both vertically and horizontally,

$$A_{p-1} = \begin{pmatrix} A_{p-1}^{hs,hs} & A_{p-1}^{hs,ls} & A_{p-1}^{hs,hf} & A_{p-1}^{hs,lf} \\ A_{p-1}^{ls,hs} & A_{p-1}^{ls,ls} & A_{p-1}^{ls,hf} & A_{p-1}^{ls,lf} \\ A_{p-1}^{hf,hs} & A_{p-1}^{hf,ls} & A_{p-1}^{hf,hf} & A_{p-1}^{hf,lf} \\ A_{p-1}^{lf,hs} & A_{p-1}^{lf,ls} & A_{p-1}^{lf,hf} & A_{p-1}^{lf,lf} \end{pmatrix},$$

where $A_{p-1}^{i,j}$ is $n^i \times n^j$ for $i, j = hs, ls, hf, lf$, then we can write

$$\tilde{A} = \begin{pmatrix} A_{p-1}^{hs,hs} & 0 & A_{p-1}^{hs,ls} & 0 & A_{p-1}^{hs,hf} & A_{p-1}^{hs,lf} & I & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & I & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ A_{p-1}^{ls,hs} & 0 & A_{p-1}^{ls,ls} & 0 & A_{p-1}^{ls,hf} & A_{p-1}^{ls,lf} & 0 & I & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 & I & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ A_{p-1}^{hf,hs} & 0 & A_{p-1}^{hf,ls} & 0 & A_{p-1}^{hf,hf} & A_{p-1}^{hf,lf} & 0 & 0 & I & 0 & 0 & \dots & 0 \\ A_{p-1}^{lf,hs} & 0 & A_{p-1}^{lf,ls} & 0 & A_{p-1}^{lf,hf} & A_{p-1}^{lf,lf} & 0 & 0 & 0 & I & 0 & \dots & 0 \\ A_{p-2}^{hs} & 0 & A_{p-2}^{ls} & 0 & A_{p-2}^{hf} & A_{p-2}^{lf} & 0 & 0 & 0 & 0 & I & \ddots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ A_1^{hs} & 0 & A_1^{ls} & 0 & A_1^{hf} & A_1^{lf} & 0 & 0 & 0 & 0 & 0 & \dots & I \\ A_0^{hs} & 0 & A_0^{ls} & 0 & A_0^{hf} & A_0^{lf} & 0 & 0 & 0 & 0 & 0 & \dots & 0 \end{pmatrix}.$$

The first, third, fifth and sixth block rows of this matrix contain n^{hf}, n^{lf}, n^{hf} and n^{lf} rows respectively. The second and fourth block rows (columns) contain n^{hf} and n^{lf} rows (columns) respectively and the identity matrices in block columns 7–10 identify them as containing n^{hs}, n^{ls}, n^{hf}, and n^{lf} columns respectively. Outside the first six rows and ten columns the blocks are $n \times n$.

This matrix is similar to the transition matrix in equation (8) of Thornton and Chambers (2017) apart from no longer dividing the flow variables by the time span and a reordering carried out by pre-multiplication by (and post-multiplication by the transpose of) the permutation matrix

$$P = \begin{bmatrix} 0 & 0 & I_{n^hs} & 0 & 0 \\ 0 & 0 & 0 & I_{n^{ls}} & 0 \\ I_{n^{hf}} & 0 & 0 & 0 & 0 \\ 0 & I_{n^{lf}} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I_b \end{bmatrix}.$$

As P is orthogonal $\tilde{C}$ is similarly a reordering of the exponential matrix in equation (11) of Thornton and Chambers (2017), without dividing the flows by the time span.

Lemmas In moving through the equations (4) to (12) we will make use of the following easily verified results concerning the $m \times n$ matrix $A = [a_1, a_2, \dots a_n]$, with j'th column, $a_j = [0, 0, \dots, 0]'$ and $n \times n$ matrix $B = [b_1, b_2, \dots b_n]$, with j'th column, $b_j = [0, \dots, 0, 1, 0, \dots, 0]'$ with the 1 in the $j'th$ cell.

Lemma A1) The j'th column of CA equals a_j, for any conformable matrix C .

Lemma A2) The j'th column of B^k equals b_j, for any $k = 0, 1, 2, 3, \ldots$.

Lemma A3) The j'th column of CAB^k equals a_j, for any $k = 0, 1, 2, 3, \ldots$.